

G06F 17/30

[21] 申请号 01114735.0

**[11]公开号 CN 1387140A**

段修明 罗才洋

[illegible]

ISSN 1008-4274

1. 一种线上资料撷取分析的代理服务系统，其特征在于：该系统包括有：原始数据库，它内存贮有原始资料；代理服务器，具有控制分析功能的模块；客户端，包括有发送及接收讯息的功能模块；利用该系统，可以通过客户端向代理服务器发出请求，由代理服务器根据请求从原始数据库中撷取原始资料并予以分析，并将分析结果传送至客户端。

2. 根据权利要求1所述的线上资料撷取分析的代理服务系统，其特征在于：代理服务器上控制分析功能的模块包括有控制处理模块、本地数据库及数据分析模块。

3. 一种线上资料撷取分析的代理服务方法，可使代理服务器根据客户从客户端发出的请求而从远程的原始数据库中自动撷取所需原始资料，并对这种原始资料予以分析，其特征在于：它包括以下步骤：

(a) 客户于客户端向代理服务器发出资料分析的请求讯息；

(b) 代理服务器接收到请求讯息后，以该请求讯息的内容作为查询条件进行原始数据库的查询；

(c) 代理服务器对原始数据库中符合查询条件的若干原始资料进行适当处理得到若干资料记录，并将这种资料记录写入代理服务器所包括的本地数据库中；

(d) 代理服务器对存贮于本地数据库的资料记录进行分析处理，并将分析的结果自动生成一分析报告；

(e) 代理服务器将分析报告传送至客户。

4. 根据权利要求3所述的线上资料撷取分析的代理服务方法，其特征在于：代理服务器是通过访问一网页服务器来查询原始数据库的原始资料。

5. 根据权利要求3所述的线上资料撷取分析的代理服务方法，其特征在于：前述符合查询条件的若干笔原始资料是分别借由若干超文本所链接；这种超文本设置于一查询结果网页上，而超文本所

链接的原始资料也以一原始资料网页形式展示。

6. 根据权利要求3所述的线上资料撷取分析的代理服务方法，其特征在于：代理服务器又包括有一控制处理模块，代理服务器通过该控制处理模块对原始数据库中符合查询条件的若干原始资料进行适当处理，以得到若干笔资料记录。

7. 根据权利要求3所述的线上资料撷取分析的代理服务方法，其特征在于：代理服务器又包括有一数据分析模块，代理服务器是通过该数据分析模块对存贮于本地数据库的若干笔资料记录进行分析处理。

8. 根据权利要求3所述的线上资料撷取分析的代理服务方法，其特征在于：步骤(c)又包含以下过程：

(c-1) 获得上述查询结果网页的原始码；

(c-2) 找到超文本所对应的原始码，以得到该超文本所对应的链接地址；

(c-3) 依照上述超文本所对应的链接地址，控制处理模块打开相对应的原始数据网页；

(c-4) 依据代理服务器的本地数据库上预定字段名称在于原始资料网页的原始码中查找相应字段名称的字符，并将字段名称所对应的资料写入本地数据库中；

(c-5) 在查询结果网页的原始码中寻找下一条超文本，如果有，则回到步骤(c-2)重复执行，如果没有，则结束。

9. 一种线上资料撷取分析的代理服务的交易方法，它可使客户向服务供应商发出资料分析的请求，由服务供应商所具有的代理服务器代理它完成向原始数据站点下载相关资料并作分析，最后在线上完成交易，其特征在于：该方法包括以下步骤：

(a) 客户于客户端向代理服务器发出资料分析的请求讯息；

(b) 代理服务器接收到请求讯息后，以该请求讯息的内容作为查询条件进行原始数据库的查询；

(c) 代理服务器根据原始数据库中符合查询条件的若干原始资料数量计算出所需服务费用；

(d) 代理服务器将符合查询条件的若干原始数据清单及所需费用讯息发送给客户，以让客户进行交易确认；

(e) 代理服务器在接收到客户的交易确认讯息之后，对原始数据库中符合查询条件的若干原始资料进行适当处理得到若干资料记录；并将这种资料记录写入代理服务器所包括的本地数据库中；

(f) 代理服务器对存贮于本地数据库的资料记录进行分析处理，并将分析的结果自动生成一分析报告；

(g) 代理服务器将分析报告传送至客户；

(h) 客户支付根据步骤(c)中所计算的服务费用给服务供应商。

## 线上资料撷取分析的代理服务系统及方法

本发明是有关于一种可以线上撷取并分析资料的系统及方法，特别是指一种可以根据客户要求而提供线上撷取并分析资料的代理服务的系统及方法。

在知识经济到来的时代，"知识" (knowledge) 越来越受到广泛的重视。"知识" 是一种流动性质的综合体，其中包括结构化的经验、价值、以及经过文字化的信息 (information) 等。而所述的信息是将来源于报章、杂志及网站资料来源等这些未经整理分析过的原始资料 (data) 经过诸如分类、整合及分析后所得到的，可以让看的人了解其中的意义。如何将数量巨大的资料准确而快速地转化为有用的信息，或是更深层的知识，正日益成为业界广泛关注的课题。

不断普及与发展的网际网络特别是国际互联网 (Internet) 已成为一种查询及获取相关资料的重要工具，用国际互联网可快速查询并获取全面的各种资料，然而，国际互联网上的资料种类繁多而又数量巨大，使用者很难从浩如烟海的资料中找到它所需要的资料，为帮助使用者用最有效率的方式，找到切中需求的资料，可以借由现有的关键词查询检索工具及搜索引擎技术从国际互联网上萃取过滤资料，但即使如此，最终得到的查询结果数量仍很巨大，要从数百，甚至上千笔资料中的统计分析中得到相关的分析结果，对于不是资料分析专业人员的普通用户来讲并非易事，而且，由于目前的网络频宽不足的限制，透过网络下载大量的资料，会存在下载速度慢，耗费用户较多时间的问题，且较容易连网断线而使资料下载中断，使用户无法及时得到所需资料。

请参见图1所示，它是一种现有的为客户提供资料分析的服务模式，其中客户 (未图出) 需要在一客户端3的计算机上透过国际互联网向远程一原始数据库10 (亦可为一存放原始资料的网站) 查

询出它所需要的资料，并从客户端3下载这种资料，由于用户所查询到均为一些未经处理过的原始资料，而且这种原始资料往往数量较为巨大，依靠人工的方式需花较多时间方能对上述的大量原始资料做到正确的处理与统计，故相关的软件供应商2专门为客户提供对特定的原始资料做统计分析的软件系统，例如AuriginTM公司的软件产品IPAM System（相关网址为：<http://www.aurigin.com>），以及Wisdomain公司的专利资料分析软件产品PatentLabTM-II（相关网 址 为：<http://www.wisdomain.com>、<http://www.delphion.com>）。当客户需要分析原始资料时，须事先在相关的软件供应商2处购买或免费下载相关的统计分析软件，然后安装至客户端3的计算机上后，再利用该软件系统对原始资料做相应的分析，它为一种典型的客户端/服务器（Client/Server）模式的服务方法。综上，现有的资料分析模式通常须经过以下三个步骤：

- (i) 客户从服务供应商处购买（或免费下载）资料分析软件；
- (ii) 购买并下载原始资料文件；
- (iii) 客户自行用资料分析软件对原始资料做分析。

上述现有的资料分析方式的不足在于：首先，客户进行资料分析的最终需求及目的各不相同，某些客户只关心资料分析的结果或结论，而不关心资料分析的过程，或者，也许客户并不是资料分析的专业人员，不知道如何查询它所想得到的原始资料，也不了解如何分析资料方能得到它想要的结果，而现有的资料分析的动作是由客户自己完成，它可能无法正确的进行资料分析，或不能从原始资料中得到它想要的正确分析结果；其次，随着资料分析技术的不断发展，同时客户对于资料分析的数量，以及资料分析的深度要求愈来愈高，客户从服务供应商处购买（或免费下载）的资料分析软件在经过一定时间后可能会有新的功能增加或改进，而客户必须对该软件不断升级，并可能会有相应的费用产生，故现有的资料分析服务模式不能动态而实时满足客户的更高要求；再次，如前所述，客户须将大量原始资料购买至本地后才能开始资料分析，当资料数量

很巨大时，客户要花较长时间下载这种资料，并需用占用较大的存储设备容量，但是，有时客户仅仅需要对这种原始资料分析后的结果，而资料分析完成后这种原始资料并无更多的利用价值，所以，现有的资料分析的服务模式会有浪费客户资源（时间、费用等）的现象。

本发明目的在于提供一种线上资料撷取分析的代理服务系统及方法，它可由客户通过网际网络向服务供应商发出资料查询及分析请求，而由服务供应商代理客户进行资料撷取并分析，最后将分析结果传送给客户。

为达到上述目的，本发明采用了以下技术方案：本发明提供一种线上资料撷取分析的代理服务的系统与方法，可使代理服务器根据客户从客户端发出的请求而从远程的原始数据库中自动撷取所需原始资料，并对这种原始资料予以分析。首先，客户在客户端向代理服务器发出资料分析的请求讯息；代理服务器接收到请求讯息后，以该请求讯息的内容作为查询条件进行原始数据库的查询；代理服务器对原始数据库中符合查询条件的若干原始资料进行适当处理得到若干资料记录，并将这种资料记录写入代理服务器的本地数据库中；代理服务器对存贮于本地数据库的资料记录进行分析处理，并将分析的结果自动生成一分析报告；最后代理服务器将分析报告传送至客户。

相较于现有技术，本发明提供根据客户需要而定制的资料查询、下载及整理分析，可使客户得到高效率而具附加价值的专业服务，且本发明提供方便可靠的线上交易方式，可使客户能够在线上及时得到其所需要的资料撷取分析服务。

下面结合附图及较佳实施例对本发明作进一步说明。

图1为一种现有的资料下载并分析的模式示意图。

图2为本发明线上资料撷取分析的代理服务系统的整体架构图。

图3为本发明线上资料撷取分析的代理服务系统的线上撷取分析资料过程的方法流程图。

图4为本发明线上资料撷取分析的代理服务系统的线上自动撷取资料的作业流程图。

图5为一在美国专利商标局网站用"ICL/G06F"的查询条件所得到的查询结果Web网页的局部图。

图6为图5所示Web网页的原始码片段。

图7为图5所示Web网页上超文本所链接的详细内容Web网页局部图。

图8为图7所示Web网页的原始码片段。

图9为资料撷取分析的代理服务系统的线上交易方法流程图。

为便于理解，以下对本发明所涉及的专业术语做简要的说明：

超链接（Hyperlink），指一个文档到另外一个文档的导航链接，通常，超链接会以高亮文字显示，当以鼠标点击该超链接时，则会跳到所链接的另外一个文档上。

超文本（Hypertext），一种全局性的信息机制，它将文档中的不同部分通过关键词建立超链接，使信息得以用交互方式搜索。

国际互联网（Internet），泛指由多个计算机网络相互连接而成的一个网络，它是在功能和逻辑上组成的一个大型网络，专指全球最大的、开放的、由众多网络相互连接而成的计算机网络。

网页（Web page），也称Web网页，是Internet上的一种超媒体显示页面，一般由HTML语言（详见后述）所编写，它可作为文字、图形、声音等多媒体的传播介质。

HTML（Hypertext Markup Language），一种撰写Web网页的直译式语言，HTML语言中允许文字中包含有定义字型、外观、图形及超文本链接的程序代码，并利用超链接的特性，提供使用者以一种顺序浏览Web网页的内容。

URL（Universal Resource Locator），在Internet的WWW服务程序上用于指定信息位置的表示方法。例如www.uspto.gov/index.htm表示美国专利与商标局的主页位置。

浏览器（Browser），Web服务的客户端浏览程序，可向网页服务器发送各种请求，并对从网页服务器发来的由HTML语言定义的



超文本信息和各种多媒体数据格式进行解释、显示和播放。

请参见图2所示，为本发明线上资料撷取分析的代理服务系统的整体架构图，该服务系统包括有原始数据库10（亦可为一存放有原始资料的网站）、服务供应商2及客户端3，其中在服务供应商2端设有本地数据库23，它可用于存贮从原始数据库10中撷取的原始资料。本发明线上资料撷取分析的代理服务系统的服务流程说明如下：首先由客户（未画出）于客户端3的计算机上向服务供应商2发送资料分析的请求讯息，当服务供应商2接收到该请求讯息后会自动将它做适当转换处理而生成标准格式的查询讯息，然后，服务供应商2将标准格式讯息发送至原始数据库10，并借由原始数据库10的查询引擎进行资料检索查询，进而得到符合前述标准化的查询条件的若干笔原始资料，然后对这种原始资料的文字做撷取分类并下载至本地数据库23的相关字段，由服务供应商2对该本地数据库23中的资料做相应分析处理，最后将分析结果传送给客户端3并向客户收取服务酬金。

如上所述，本发明线上资料撷取分析的代理服务系统可用于为客户提供一种具附加价值的代理服务机制，代理客户进行线上的资料查询、下载及分析。

请参见图3所示，是本发明线上资料撷取分析的代理服务系统的资料查询、下载及分析的系统方框图，包括有原始资料站点1、服务供应商2及客户端3，其中原始资料站点1具有一网页服务器11及一原始数据库10，原始数据库10中存贮有大量的原始资料，这些原始资料可应外界的查询请求而以web网页（以HTML语言编写）的形式通过网页服务器11向外界展示，它外界任何具有网页浏览器功能的计算机均可访问该网页伺服器11并查询到所需原始数据。服务供应商2包括有一控制处理模块21、数据分析模块22及一本地数据库23，客户端3包括有一网页浏览器31。

当原始资料站点1接收到某一资料查询条件后，会通过访问网页伺服器11得到相应查询结果网页5，于该页面上所显示的查询结果为若干个超文本50，每个超文本50均以超链接方式对应一详细内

容网页51。服务提供商2的控制处理模块21自动撷取上述这种查询结果讯息，并计算资料总数及所需服务酬金数目，通知客户端3以确认购买（详见后述）。

当收到客户的确认购买讯息后，控制处理模块21即开始自动下载并剖析查询结果所对应的详细内容网页51，请参见图4所示，为资料自动下载剖析的流程图，它包括的步骤如下所述，为使该流程的说明更易于理解，结合一具体实例说明：

- (a) 根据查询条件得到查询结果的Web网页5；请参见第五图所示，为一于美国专利商标局（相关网址：<http://www.uspto.gov>）用"ICL/G06F"的查询条件所得到的查询结果Web网页5的局部图，其中所查询的符合条件的专利是以超文本50的方式显示，这种超文本50是以超链接的方式与各具体专利说明书内容对应。
- (b) 获得Web网页5的HTML原始码（source code）；请参见图6所示，为图5的Web网页5的HTML原始码片段。
- (c) 找到超文本50所对应的HTML原始码，并得到该超文本所对应的URL；请参见图6的HTML原始码，其中< A HREF = "http://patents.uspto.gov/cgi-bin/ifetch4?ENG+PATBIB-1999-2000+0+990662+0+1+165850+F+1+19984+1+ICL%2fg06f"/>（第一行及第二行）是代表一个超文本50的超链接所对应的URL地址。
- (d) 依照上述的URL地址，控制处理模块21打开相对应的详细内容Web网页51；请参见图7所示，为上述步骤(c)中的URL所对应的详细内容Web网页51，也就是图五中的超文本50所链接的页面，而图8所示为图7的详细内容Web网页51的HTML原始码片段。
- (e) 依据预定的数据库字段名称在详细内容Web网页51的原始码中查找相应字符串，并将相关资料写入数据库中；例如本地数据库23的预定资料表字段的一为"Inventor"（发明人），则当控制处理模块21在上述详细内容Web网页51的HTML原始码中搜寻到字符串"Inventor"时，则将自动撷取该字符串之后的相应创

作人姓名的字符串至本地数据库23的相应字段内容中, 在本例中, 控制处理模块21将撷取"Goodwin; David W.", "Cohn; Robert S.", "Lowney; Paul G."以及"Rubin; Norman"四个字符串至本地数据库23的"Inventor"字段中, 依据相同的原理与步骤, 可以将该笔原始资料的其它相关字段讯息撷取至本地数据库23中, 而形成一笔资料记录6 (如图3所示)。

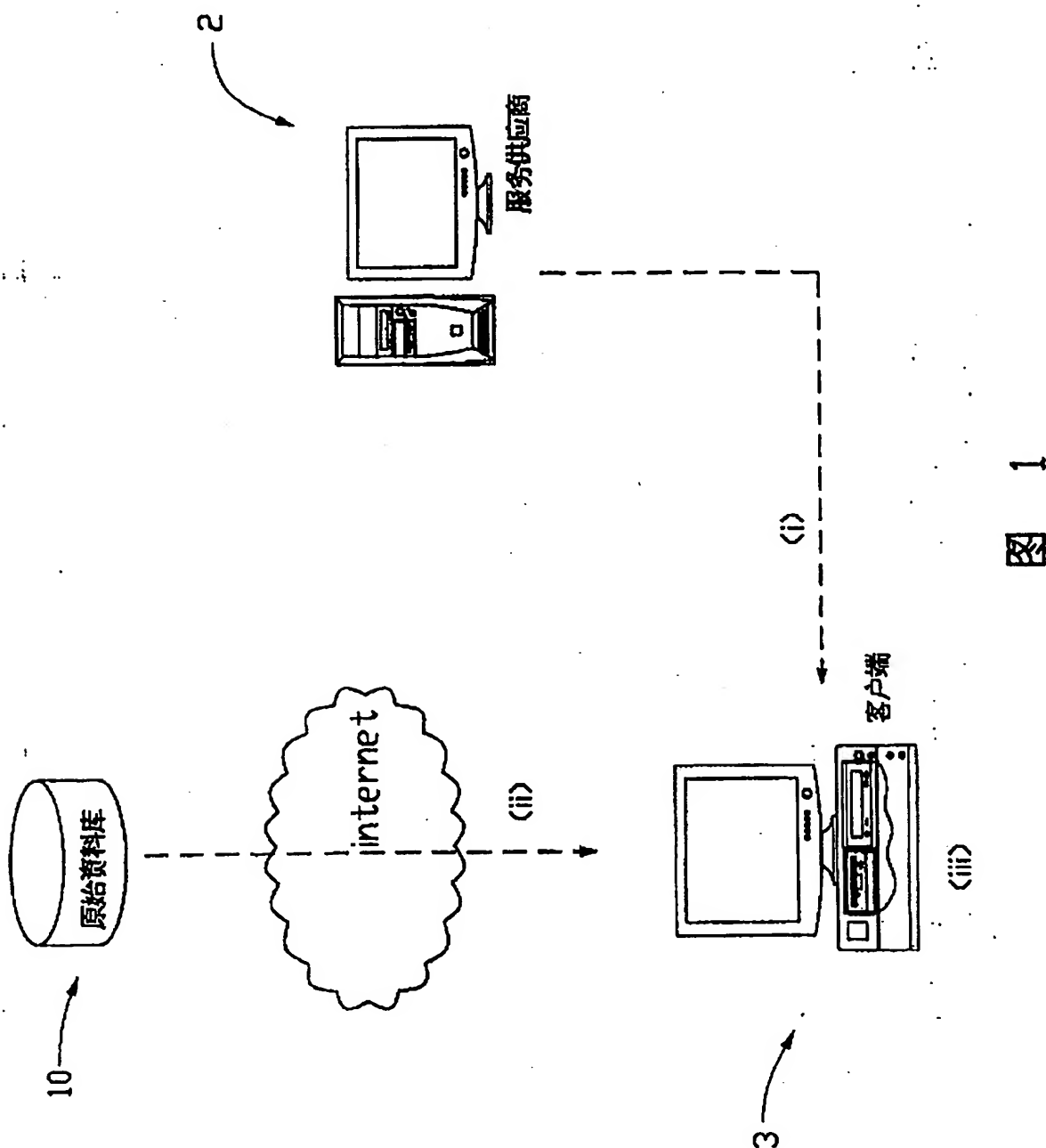
(f) 在查询结果Web网页5的HTML原始码中寻找下一条超文本50, 如果有, 则回到步骤(c), 如果没有, 则结束。利用上述方法, 可以将查询结果Web网页5的所有超文本50所链接的原始资料撷取, 转换并存贮入本地数据库23的相关栏位中, 从而产生若干笔资料记录6。

利用上述的方法, 即可实现从远程原始数据库10中查询所需原始资料, 并将这种原始数据由服务供应商2端的控制处理模块21自动撷取至本地数据库23中的效果。

请参见图9所示, 为本发明线上资料撷取分析的代理服务系统的交易过程示意图, 在该交易机制中, 包括有具网页浏览器31的客户端3、具有网页服务器11的服务供应商2以及原始资料站点1。需要资料查询分析服务的客户 (未图出) 可利用客户端3的网页浏览器31访问服务供应商2的网页服务器11, 并在输入请求Web页面40上输入请求讯息, 并将此请求讯息发送回网页服务器11; 控制处理模块21处理此请求讯息, 将它转换成标准格式的查询语句, 然后将此查询语句作为限制条件查询远程的原始资料站点1; 控制处理模块21自动将查询到的结果 (原始资料的总数及列表) 讯息及相应的费用自动生成一确认购买的Web网页41, 在该确认购买的Web网页41上具有查询结果的原始资料列表及总数, 以及根据资料总数与客户选择的分析模式计算出的费用总数, 所需费用的计算公式可为: 总费用 = (分析的资料笔数 × 单价) × 分析模式权值, 其中分析模式权值是根据资料分析所消耗服务供应商2的资源总量 (工作时间及工作量等) 经换算得到, 每种分析模式是对应于不同的资料分析内容与深度, 可供客户根据各自需要而选择, 为便理解, 现举一例说明,

在图9所示的确认购买Web网页41中共列出三种分析模式供客户选择：A、B及A+B，每种分析模式所对应的分析模式权值分别为1、1.2及1.5，如果所查询的资料笔数为200笔，每笔资料的单价为3元，客户所选择的为B分析模式，则根据上述公式计算出的总费用为： $200\text{笔} \times 3\text{元/笔} \times 1.2 = 720\text{元}$ 。客户在点选所需分析模式并确定购买后对应产生确认购买讯息。

在接收到客户的确认购买讯息后，控制处理模块21撷取所查询的原始资料并写入本地数据库23，形成若干笔资料记录6；由资料分析模块22将上述的若干笔资料记录6依据客户选择的分析模式做相应分析，并将分析结果自动产生一分析报告7；最后将此分析报告7的Web页面42向客户展示，客户在收到此分析报告7后，将根据前述的确认购买的协议向服务供应商2支付相应的费用，最后完成此次交易。



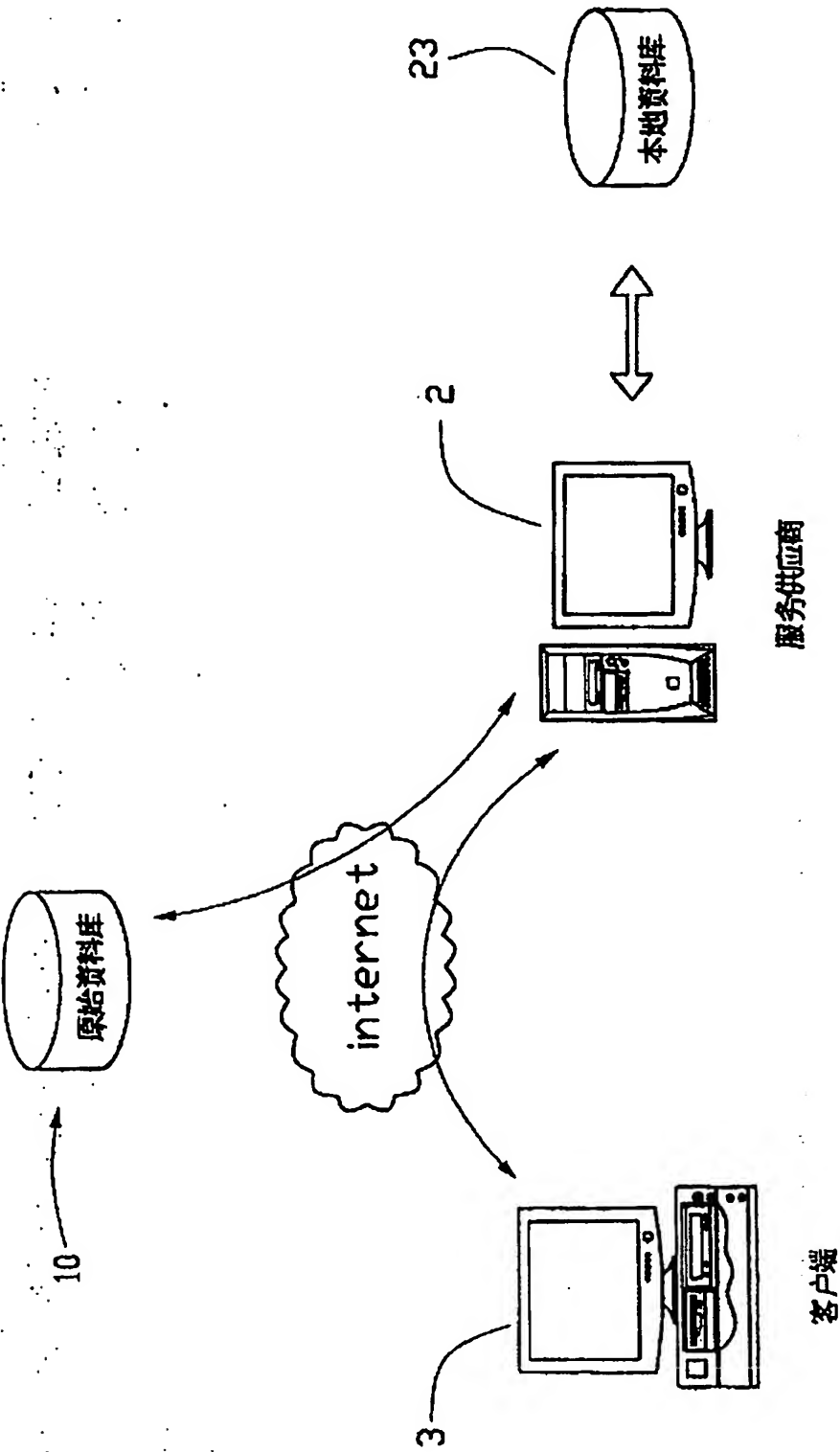


图 2



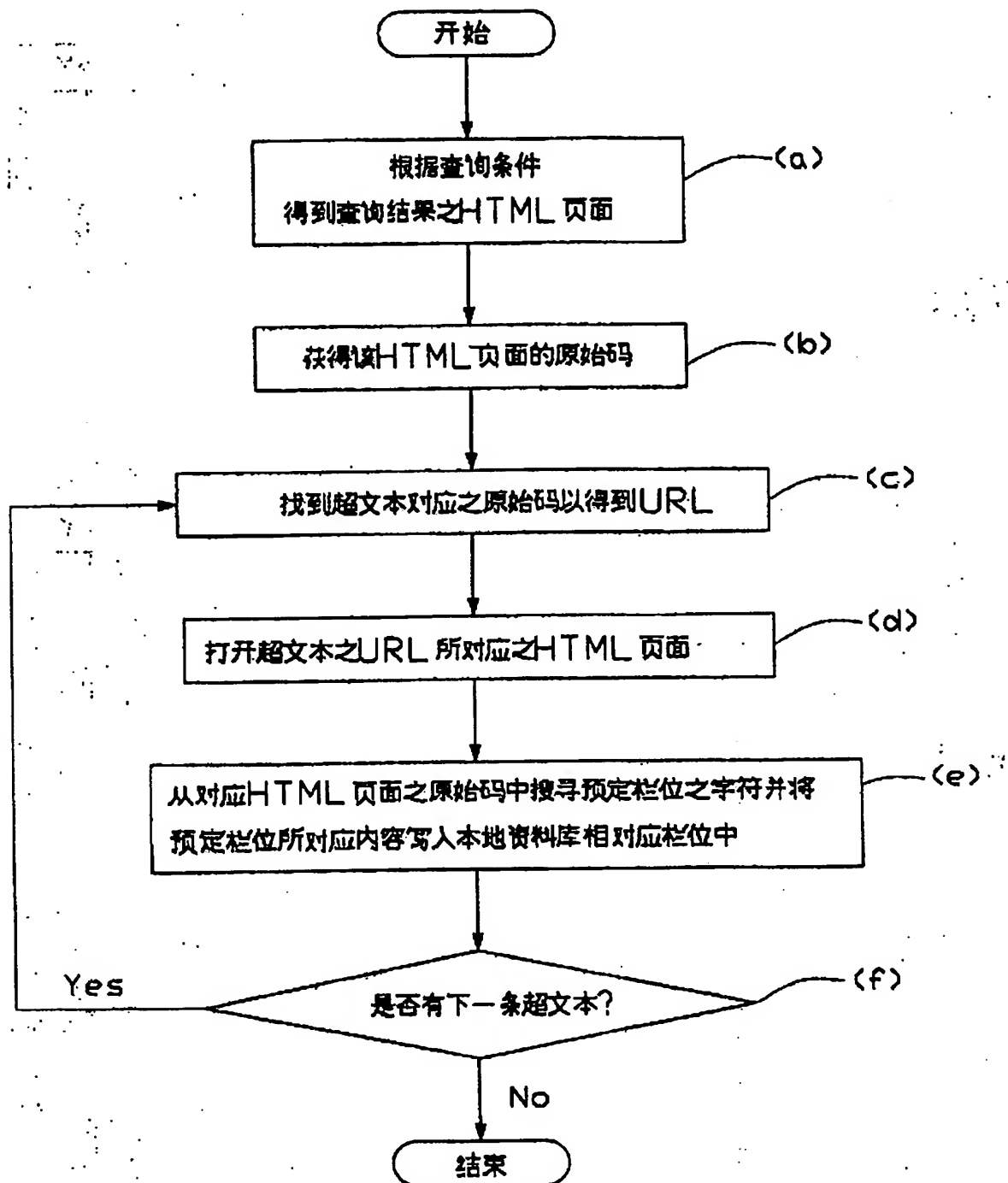


图 4



[Search Summary]

Results of search in 1999-2000 for: ICL/G06F-\*/\*:19984 patents.

Hits 1 through 50 of 19984

Refine Search

ICL/G06F

Pat. No. Title

1. 6,158,048 User transparent mechanism for profile feedback optimization
- 50 2. 6,158,048 Method for eliminating common subexpressions from java byte codes
3. 6,158,047 Client/server system for fast, user transparent and memory efficient computer language translation
4. 6,158,046 Computer device and method for processing data utilizing pseudoinstruction words
- 5 5. 6,158,045 Portable debugging services utilizing a client debugger object and a server debugger object with flexible addressing support
6. 6,158,044 Proposal based architecture system
7. 6,158,049 Signal processing apparatus and method
8. 6,158,042 Method of counting bits



5

United States Patent

6,158,049

Goodwin, et. al.

Dec. 5, 2000

User transparent mechanism for profile feedback optimization

51

## Abstract

A profile feedback optimization system is provided. The system accepts as input an original application and produces an optimized version of the application. The system first instruments the original application through an instrumentation process such that when it executes it generates profile data. During instrumentation, the system configures images in the application to trap any function calls to critical functions. This allows the instrumented version of the application to execute transparently and appear as the original version. The instrumentation process converts each original image of the application to a version that references other instrumented images. The instrumentation process also configures the operating system to execute the instrumented version instead of the original version when the user selects the application for execution. Once instrumented, the instrumented application is executed. During execution, profile data produced from the instrumented images is saved in a database. Once profile data has been generated, an optimization process is used to optimize the original version of the application based upon the profile data. The optimized version is also configured to execute just as the original application, but references optimized images instead of original images.

Inventors: Goodwin; David W. (Waltham, Ma); Cohn; Robert S. (SALEM, NH); Lowney; Paul G. (Concord, MA); Rubin; Norman (Cambridge, Ma).

Assignee: Compaq Computer Corporation().

Appl. No.: 132,449

Filed: Aug. 11, 1998

Intl. cl.:

G06F9/45



7

```

<HR>
<TABLE><TR><TD VALIGN="TOP">Inventors:</TD><TD VALIGN=TOP><B>Goodwin;David.W.</B>
(Waltham,MA);
<B>Cohn;Robert S.</B>
(Salem,NH);
<B>Lowney;Paul G.</B>
(Concord,MA);
<B>Rubin;Norman</B>
(Cambridge,MA).
</TD>
</TR>
<TR><TD VALIGN=TOP>Assignee:</TD><TD VALIGN=TOP><B>Compaq Computer Corporation</B>

```

图 8

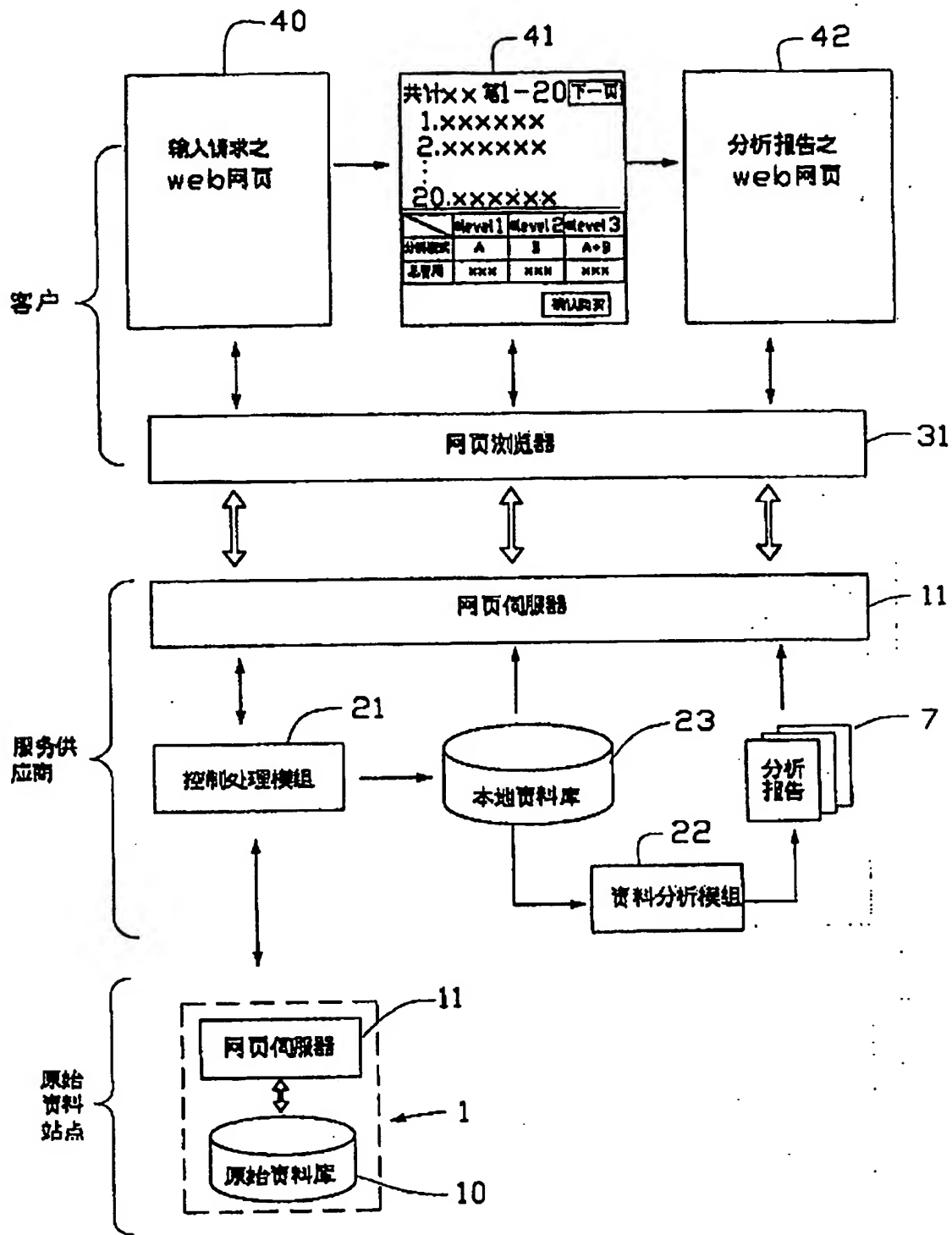


图 9

**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

**BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☐ **FADED TEXT OR DRAWING**
- ☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☒ **GRAY SCALE DOCUMENTS**
- ☐ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**